

From: Mia, Marcia
Sent: Thursday, May 30, 2019 04:01 PM
To: Marsh, Karen
Subject: FW: RE: statistics

Marcia B Mia
Air Branch
Office of Compliance
2227A WJCS
U.S. Environmental Protection Agency
202-564-7042

This message may contain deliberative, attorney-client or otherwise privileged material. Do not release this message without the appropriate review. If you are not the intended recipient, kindly advise and delete this message/attachments.

From: Schulman, Andrew
Sent: Thursday, May 30, 2019 11:59 AM
To: Mia, Marcia <Mia.Marcia@epa.gov>; Cox, John <cox.john@epa.gov>
Subject: RE: RE: statistics

Here are the results.

The mean reported survey time was 1922 s for low-production wells (n=88), and 3693 s for non-low-production wells (n=316). The mean reported number of leaks was 1.18 for low-production wells (n=82), and 3.72 for non-low-production wells (n=270). Permutation tests of equality of the means for low- and non-low-production wells, against the alternative that the low-production wells had smaller means, had p-values of $2.3 * 10^{-4}$ for the survey times, and $1.4 * 10^{-12}$ for the leak counts. These are convincing statistical evidence that the low-producing wells reported both shorter survey times and smaller leak counts, on average.

I paraphrased "number of leaks". Feel free to mark this up as you need. I can provide more detail if it's wanted, such as the method used for the test, or diagnostic plots showing that the distributions of survey times and number of leaks were really higher for the non-low-producing wells. My guess is that the above is enough for your purposes, but let me know if you want more.

Andrew E. Schulman
Office of Compliance
U.S. Environmental Protection Agency
202-564-5244

From: Mia, Marcia
Sent: Wednesday, May 29, 2019 5:44 PM

To: Schulman, Andrew <Schulman.Andrew@epa.gov>; Cox, John <cox.john@epa.gov>

Subject: RE: RE: statistics

Hi Andrew,

Thanks again for helping us with our project. I apologize I didn't get this to you COB yesterday as promised.

As we discussed, we want to know if there is a statistical difference between low production wells and non-low production wells. We are interested in two different comparisons:

1. the data in column H (survey time) between low and non-low; and
2. The data in column I (number of each component type for which fugitive emissions were detected) between low and non-low

We also discussed if the data were normal, and you noted that since the data in column I is a "count" that isn't meaningful. So whatever you think will give us some confidence in our statistical outcome is great.

I hope I cleaned up the data in columns H and I so that there are only numbers. A zero is meaningful as in it means zero, a blank means nothing reported.

Feel free to give me a call - I am working remotely tomorrow but my phone will forward - also off on Friday. << File: Copy of KRM 0000a Reports Evaluation_ERG_2019-05-15_statistics.xlsx >>

Marcia B Mia

Air Branch

Office of Compliance

2227A WJCS

U.S. Environmental Protection Agency

202-564-7042

This message may contain deliberative, attorney-client or otherwise privileged material. Do not release this message without the appropriate review. If you are not the intended recipient, kindly advise and delete this message/attachments.

From: Schulman, Andrew

Sent: Tuesday, May 28, 2019 8:59 AM

To: Mia, Marcia <Mia.Marcia@epa.gov>; Cox, John <cox.john@epa.gov>

Subject: RE: RE: statistics

Sure. I'm free til 11. Come on by any time.

Andrew E. Schulman

Office of Compliance

U.S. Environmental Protection Agency

202-564-5244

From: Mia, Marcia
Sent: Tuesday, May 28, 2019 8:55 AM
To: Cox, John <cox.john@epa.gov>; Schulman, Andrew <Schulman.Andrew@epa.gov>
Subject: RE: RE: statistics

I am running late - traffic and rain were terrible. I am free until 10:30 if you are - can come down when I get in. Should be about 945?

Marcia B Mia
Air Branch
Office of Compliance
2227A WJCS
U.S. Environmental Protection Agency
202-564-7042

This message may contain deliberative, attorney-client or otherwise privileged material. Do not release this message without the appropriate review. If you are not the intended recipient, kindly advise and delete this message/attachments.

-----Original Appointment-----

From: Mia, Marcia
Sent: Thursday, May 23, 2019 12:50 PM
To: Mia, Marcia; Cox, John; Schulman, Andrew
Subject: RE: statistics
When: Tuesday, May 28, 2019 9:30 AM-10:00 AM (UTC-05:00) Eastern Time (US & Canada).
Where: Andrew's Office

Hi Andrew,

I don't have any time this afternoon and off tomorrow. Monday a holiday. I hope this time works, feel free to counter.

Marcia B Mia
Air Branch
Office of Compliance
2227A WJCS
U.S. Environmental Protection Agency
202-564-7042

This message may contain deliberative, attorney-client or otherwise privileged material. Do not release this message without the appropriate review. If you are not the intended recipient, kindly advise and delete this message/attachments.

.....

.....

From: Mia, Marcia
Sent: Thursday, May 23, 2019 12:39 PM
To: Schulman, Andrew
Cc: Cox, John
Subject: RE: statistics
Attachments: Copy of Copy of KRM 0000a Reports Evaluation_ERG_2019-05-15_statistics.xlsx

Thank you Andrew. Here is the spreadsheet. We are still populating the "Non Low Production" worksheet. But for the purposes of showing us how to determine if the data is normally distributed (or other useful statistical functions), we can use the Low Production worksheet and the Invalid worksheet. I will set up 30 minutes to discuss. Including John Cox as he is helping.

I want to know if the data in column H and I in the Low Production Worksheet and Invalid worksheet is normally distributed. If it is, then I want to conduct a T test to determine if the average of the values is statistically the same.

Marcia B Mia

Air Branch

Office of Compliance

2227A WJCS

U.S. Environmental Protection Agency

202-564-7042

This message may contain deliberative, attorney-client or otherwise privileged material. Do not release this message without the appropriate review. If you are not the intended recipient, kindly advise and delete this message/attachments.

From: Schulman, Andrew
Sent: Thursday, May 23, 2019 11:52 AM
To: Mia, Marcia <Mia.Marcia@epa.gov>
Subject: RE: statistics

Hi Marcia. Mike forwarded your question to me. I can help. Why don't you send me your spreadsheet, and schedule half an hour to look at it. Any open time on my calendar is fine. I'm in the office today, teleworking tomorrow.

Andrew E. Schulman

Office of Compliance

U.S. Environmental Protection Agency

202-564-5244

From: Barrette, Michael

Sent: Thursday, May 23, 2019 11:09 AM

To: Schulman, Andrew <Schulman.Andrew@epa.gov>

Subject: FW: statistics

Hi Andrew,

Do you have the bandwidth to assist Marcia on this? If so, please reach out directly. Thanks.

Mike

From: Mia, Marcia

Sent: Thursday, May 23, 2019 10:00 AM

To: Barrette, Michael <Barrette.Michael@epa.gov>

Subject: statistics

Do you have anyone in your group that might help me run some statistics on some data in an excel spreadsheet? I have run the "Descriptive Statistics" for the data but I need to figure out if the data is normally distributed. I can't figure out how to do that, or if one of the values I am getting shows me that information. Thanks.

Marcia B Mia

Air Branch

Office of Compliance

2227A WJCS

U.S. Environmental Protection Agency

202-564-7042

This message may contain deliberative, attorney-client or otherwise privileged material. Do not release this message without the appropriate review. If you are not the intended recipient, kindly advise and delete this message/attachments.

